Spoken Words: Activating Text-To-Speech through Eye Closure

Fabian Hemmert
Deutsche Telekom Laboratories
Ernst-Reuter-Platz 7
10587 Berlin, Germany
mail@fabianhemmert.de

Danijela Djokic
Potsdam University of
Applied Sciences
Pappelallee 8-9
14469 Potsdam, Germany
djokic@fh-potsdam.de

Reto Wettach
Potsdam University of
Applied Sciences
Pappelallee 8-9
14469 Potsdam, Germany
wettach@fh-potsdam.de

CHI 2007, April 28 – May 3, 2007, San Jose, USA
ACM 1-xxxxxxxxxxxxxxxxxx.

Abstract
Writing is a predominantly visual task. In the following, we describe a system that adds a new modality to computer-supported text editing: Listening with closed eyes.

We present a prototype, for which we combined a text-to-speech software, Microsoft Word and a face tracking system. It reads out the current sentence in Microsoft Word as soon as the user closes both eyes.

In this paper we discuss further details of the concept and its background, describe the implementation of the prototype and disclose first user testing results, before we conclude our findings with an outlook on future research.

Keywords
Audiovisual, eye closure, face tracking, multimodal interfaces, text editing

ACM Classification Keywords
H5.2 [Information interfaces and presentation]: User Interfaces - Input devices and strategies

Introduction

While every writer, be it at the computer or with pencil and paper, has his own style of improving his work, repeated reading and henceforth editing the text is common practice among authors of all skill levels.

Text editing itself is a highly visual task, and especially for on-screen work, several health issues have been discussed, among them being eye-strain problems that result from lowered blink rate [1].

Fig. 1: Visual text editing in Microsoft Word.

A commonly recommended relief for these issues is to simply close the eyes, and take frequent breaks [2]. In this study, we investigated how closed eyes could not only serve as a kind of eye-strain relief, but also be valuable in terms of *productivity*.

The proposed system allows the user to listen to a text he is editing in Microsoft Word, as soon as he closes his eyes. We hypothesized that this feature would indeed not only have relaxation-related advantages: It could offer its users a *perspective change*, which we will discuss in the following.

Background

The human eyelid serves a variety of purposes, such as the protection and moisturization of the eyeball [3], and the prevention of light from entering through the pupil [4]. Closing the eyes is important for humans during the onset phase of sleep [5], and is a central element of prayer and meditation in many cultures of the world [6][7][8].

In the context of GUIs, on the other hand, closing the eyes has never been of much interest: The prevalent modality of the *graphical* user interface is, of course, the visual. Here, closing the eyes seems just counterproductive.

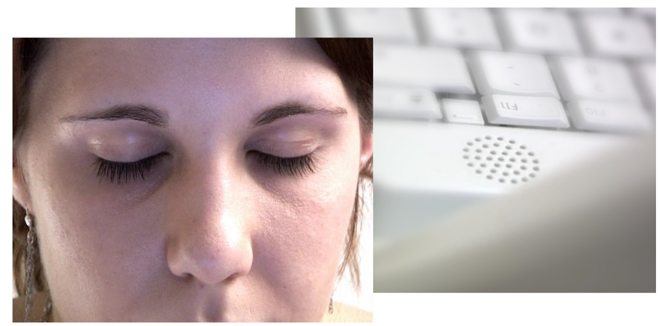

Fig. 2: Auditive rehearsal upon eye-closure. The current sentence is read out loud through a text-to-speech software

From a psychological point of view, removing a sense causes our body to balance this lack through the other senses [9]. For instance, it has been shown that people who lost their sight at an early age or at birth were able to determine pitch differences in tones significantly better than sighted people [10].

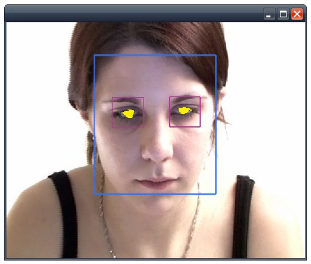

Fig. 3a: Eyes-open state of the eyelid tracking software

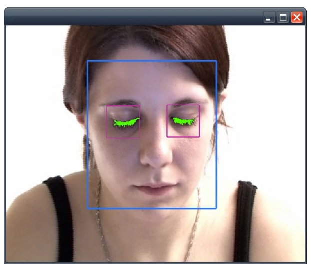

Fig. 3b: Eyes-closed state of the eyelid-tracking software, distinguishing pupils from eyelashes.

Related Work

Alternative modalities have been an active field for researchers and designers in the past. Some of these approaches replaced the visual channel [11][12] with others (e. g. to make computers accessible for the impaired people), others combined several modalities into *multimodal* ways of interaction [13][14].

While non-visual modalities have always been of interest in these investigations, assessing the eye-closure of sighted users to initialize a *non-visual mode* has yet to be explored.

Obviously, eye-closure has already been a possible user input since the first eye tracking applications were developed [15][16][17]. However, these were as well rather targeted to bodily impaired people, and it has often been argued that the eye is a sensory organ, and not a point-and-click device.

It has been argued that proprioceptive feedback is a helpful means for the avoidance of *mode errors* [18], and closing the eyes would indeed add a strong proprioceptive feedback to its very mode.

We believed that we had found a way of interacting with closed eyes in a beneficial and natural way. It was the goal of this study to determine how helpful the system would prove itself in practice.

Implementation

The prototype was implemented using a commercially available text-to-speech product, Natural Readers [19], Microsoft Word [20], a face-finding software [21], a web cam, and a Java program that was able to create virtual keyboard events.

The software was run on a 2.0GHz dual-core laptop, with the web cam placed on the upper edge of the screen. The face tracking operated at ~50fps, with a delay of ~250ms.

Given the image data of every eye, we were able to determine an open eye from a closed one by detecting the pupils inside the eyes and distinguishing them from a closed eyelid through a analysis of its shape and aspect ratio (Fig. 3a+b).

It is possible to mark the sentence at the cursor position in Microsoft Word by pressing the *F8* three times in succession, so it can be copied (*Ctrl-C*) to the clipboard. Furthermore assigning *Ctrl-F9* as the global shortcut for the text-to-speech software to speak the contents of the clipboard made it possible to have the current sentence read out loud through a sequence of automated keystrokes.

Leaving the normal text editing capabilities of Word untouched, the eyelid tracker waited for the user to close his eyes for more than 0.5 seconds. Once the closure was detected, the Java software generated the keystroke sequence as mentioned above and the sentence was read out.

User Test

The proposed system was tested by a group of volunteer users (3f, 6m, Ø25.7yrs.) of different skill levels. A control group (3f, 6m, Ø24.8yrs.), also including experts and novice users, was able to activate the text-to-speech functionality through a usual GUI button, and *not* instructed to close their eyes. Both groups were not informed about the expected outcomes of the experiments in advance.

The users were given no particular task, but asked to try out the system after a short introduction. A questionnaire was handed to them afterwards, in which they were asked to rate the subjective *bodily ergonomics* and *cognitive ergonomics* of the system, on 5 point scales.

The bodily ergonomics scale reached from exhausting (-2) to relaxing (+2), the cognitive ergonomics scale reached from impairing (-2) to helpful (+2). Both scales included a neutral answer (0).

Test Results

A Mann-Whitney test was conducted on the results of the questionnaire. The *bodily ergonomics* were rated to be *more relaxing* in the experimental group (U = 16,000; p = .024). No significant differences between the groups were found for the *cognitive ergonomics*, however we were able to determine a tendency that the eye-closure controlled system would be rated as *more helpful* than its mouse-controlled equivalent (U = 19,500; p = .058) (Fig. 4).

The user comments in the interviews confirmed this: In the experimental group, users remarked that the feature was "helpful", "relaxing" and offered a "interesting feeling to hear what the text *sounded* like", but also that it would be "more time consuming" than simply reading the text again.

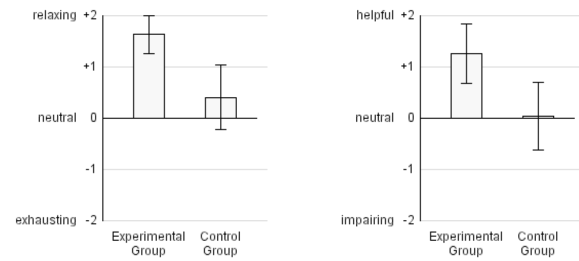

Fig. 4: Subjective bodily (l.) and cognitive (r.) ergonomics for the proposed application; significantly higher ratings for bodily ergonomics (relaxation) in the experimental group, no significant differences between the groups for cognitive ergonomics (p = .058)

Further remarks included that "spelling errors (*w*ould) probably not be found through listening". On the other hand, "errors could be found which had previously been *overseen"*. Users also hypothesized that they might perceive the text differently, "possibly also in a more objective way", because it was "read out by a another person's voice", compared to when they read it themselves.

Discussion

The proposed system has demonstrated some strengths compared to the GUI-based version in the experimental group. Combining eye-closure with text-to-speech functionality adds further benefit to closing the eyes, which, in the computer context, only had relaxation advantages to date.

It was criticized by the users that reading the text normally would be much faster, and that closing the eyes might be misinterpreted by the tracking system.

Nonetheless, the overall ratings of the feature were better in the experimental group than in the control group. Furthermore, we conclude from the users' statements that closing the eyes helped them to *focus on listening*.

The users stated that, on the one hand, they were able to find some errors in their work they *oversaw* before, but on the other hand, they *overheard* errors they might have noted while reading. This makes clear that the proposed system is not a replacement for re-reading the text, but a helpful addition.

Conclusion

We believed that bypassing visuality in software interfaces could be advantageous, and “spoken words”, as opposed to “written words”, seem to be a promising approach.

The system was well accepted among the test users, and showed facilities, compared to the GUI-based “eyes-open” alternative. Paradoxically, the users saw *less*, but the functionality *increased* their ratings of the experience.

Outlook

The proposed system offers the user a perspective change on his text. Perspective changes, be they physical or mental, are essential for our understanding of various matters, and we argue that a change away from the visual perspective is a interesting one, especially in the computer context.

With regards to this particular application, a long term study, distinguishing between professional and amateur writers could offer insights into the actual trade-offs between time-on-task, eye relaxation and the newly gained “eyes-closed perspective”.

While removing a sensory channel to improve the result sounds inherently paradoxical, we encourage further research in this field, especially with regards to eyes-closed interaction in the HCI field.

References

1. Ergonomic Problems of Eye Strain, http://www.safecomputingtips.com/ergonomic-problems-eye.html
2. Dealing With Eyestrain, http://www.upmc.com/HealthManagement/ManagingYourHealth/PersonalHealth/Men/?chunkiid=14167
3. Evinger C, Manning KA, Sibony PA, Eyelid Movements: Mechanisms and Normal Data. *Investigative Ophthalmology and Visual Science* (32), pp. 387-400, 1991
4. Meronk F, Atlas of Eyelid and Facial Anatomy, http://www.drmeronk.com/blephnews/anatomy.html
5. Davis et al., Human Brain Potentials During the Onset of Sleep. *Journal of Neurophysiology* (1), pp. 24-38, 1938
6. *Buddhism in a Nutshell: Anapana Sati*, http://www.accesstoinsight.org/lib/authors/narada/nutshell.html
7. *Ways of Christ: Christian Meditation*, http://www.ways-of-christ.net/topics/meditation.htm
8. *Shema Israel: Jewish Prayer and Ritual*, http://www.jewishmag.com/49mag/shema/shema.htm
9. Tresniowski A, Arias R, The Boy Who Sees With Sound. *People Magazine* (6), pp. , 2006.
10. Gougoux et al., Pitch discrimination in the early blind. *Nature* (430), p. 309, 2004.

11. Brewster et al., Multimodal 'eyes-free' interaction techniques for wearable devices, In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 473-480, 2003.

12. Bellik, Y. and Burger, D. Multimodal interfaces: new solutions to the problem of computer accessibilty for the blind. In *Conference Companion on Human Factors in Computing Systems*, pp. 267-268, 1994.

13. Opera Multimodal Browser, http://www.opera.com/products/devices/multimodal/

14. *Multimodal Web Applications with XHTML+Voice*, http://www-306.ibm.com/software/pervasive/multimodal/

15. Tobii Technology AB, http://www.tobii.com/

16. SR Research Ltd., http://www.sr-research.com/

17. Morimoto CH, Koons D, Amir A, Flickner M, Pupil detection and tracking using multiple light sources. *Image and Vision Computing* (18), pp. 331-335, 2000.

18. Rszkin A, Visual Feedback: Why Modes Kill, http://www.humanized.com/weblog/2006/12/07/is_visual_feedback_enough_why_modes_kill/

19. NaturalSoft Ltd., Natural Reader, http://www.naturalreaders.com/

20. Microsoft Corporation, Microsoft Word, http://office.microsoft.com/en-us/word/default.aspx

21. Froeba B, Kueblbeck C, Face Detection and Tracking using Edge Orientation Information, *SPIE Visual Communications and Image Processing* (4310), pp. 583-594, 2001.